ANNALES
POLONICI MATHEMATICI
XXX (1974)

A functional equation and its application to information theory

by P. KANNAPPAN and C. T. NG (Waterloo, Ontario)

Abstract. Let $P = (p_1, p_2, \ldots, p_n)$, $Q = (q_1, q_2, \ldots, q_n)$, $R = (r_1, r_2, \ldots, r_n)$ be three complete probability distributions, that is $p_i, q_i, r_i \geqslant 0$ with $\sum p_i = \sum q_i = \sum r_i = 1$. Then the measure generalized directed divergence is defined by

$$I_n(P||Q|R) = \sum_{i=1}^{n} p_i \log(q_i/r_i). \tag{1}$$

This measure has been characterized by different sets of postulates. While characterizing (1), one comes across the functional equation

$$f(x, y, z) + (1-x) f\left(\frac{u}{1-x}, \frac{v}{1-y}, \frac{w}{1-z}\right) = f(u, v, w) + (1-u) f\left(\frac{x}{1-u}, \frac{y}{1-v}, \frac{z}{1-w}\right), \tag{2}$$

which has been solved for f under the assumption of f having continuous first partial derivatives. In this paper, we treat (2) under the weaker assumption of the Lebesgue measurability of f in each of its variables on a proper domain and prove that f has the form,

$$f(x, y, z) = -a[x\log x + (1-x)\log(1-x)] + \left(b\log\frac{y}{1-y} + c\log\frac{z}{1-z} + d\right)x + b\log(1-y) + c\log(1-z),$$

where a, b, c, d are arbitrary constants.

1. Introduction. Let $P = (p_1, p_2, \ldots, p_n)$, $Q = (q_1, q_2, \ldots, q_n)$ and $R = (r_1, r_2, \ldots, r_n)$ be discrete probability distributions with $p_i, q_i, r_i \geqslant 0$, $\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i = \sum_{i=1}^{n} r_i = 1$. A measure generalized directed divergence is defined as ([1], [4], [6])

$$I_n(P||Q|R) = \sum_{i=1}^{n} p_i \log\frac{q_i}{r_i}. \tag{1}$$

This measure has been characterized by different sets of postulates ([2], [4]). While characterizing (1), one comes across the functional equation

$$(2)\quad f(x,y,z)+(1-x)f\left(\frac{u}{1-x},\frac{v}{1-y},\frac{w}{1-z}\right) = f(u,v,w)+(1-u)f\left(\frac{x}{1-u},\frac{y}{1-v},\frac{z}{1-w}\right)$$

which has been solved under the assumption that f possesses continuous first partial derivatives ([4]). In this paper we describe the solution of (2) under the weaker assumption of Lebesgue measurability of f in each of its variables.

We shall denote by $I=[0,1]$, $I^0=]0,1[$ and $I_1=[0,1[$.

2. A functional equation and its measurable solutions.

THEOREM. *Let $F: I\times I^0\times I^0\to R$ be a real-valued function, Lebesgue measurable in each of its three variables, satisfying the functional equation*

$$(3)\quad F(x,y,z)+(1-x)F\left(\frac{u}{1-x},\frac{v}{1-y},\frac{w}{1-z}\right) = F(u,v,w)+(1-u)F\left(\frac{x}{1-u},\frac{y}{1-v},\frac{z}{1-w}\right)$$

for all $x,u\in I_1$, $y,z,v,w\in I^0$ with $x+u\in I$, $y+v$ and $z+w\in I^0$. Then F is given by

$$(4)\quad F(x,y,z) = AS(x)+\left[d_1\log\frac{y}{1-y}+e_1\log\frac{z}{1-z}+e_4\right]x+ + d_1\log(1-y)+e_1\log(1-z),$$

where A, d_1, e_1 and e_4 are constants, S is the Shannon function given by $S(x)=-x\log x-(1-x)\log(1-x)$. The converse is also true.

Proof. We shall prove this theorem through five lemmas. We make use of the following result we gave in [3]:

PROPOSITION. *Let $f,h\colon I_1\to R$ and $g,k: I\to R$ be Lebesgue measurable functions such that*

$$(5)\quad f(x)+(1-x)g\left(\frac{y}{1-x}\right) = h(y)+(1-y)k\left(\frac{x}{1-y}\right)$$

holds for all $x,y\in I_1$, with $x+y\in I$. Then

$$(6)\quad \begin{aligned} f(x) &= AS(x)+B_1x+D,\\ g(y) &= AS(y)+B_2y+B_1-B_4,\\ h(x) &= AS(x)+B_3x+B_1+B_2-B_3-B_4+D,\\ k(y) &= AS(y)+B_4y+B_3-B_2 \end{aligned}$$

for $x \in I_1$ and $y \in I$, where S is the Shannon function and A, B_i, D are real constants.

For each fixed y, z, v and w, (3) is of the form (5). Hence there exist constants $A(y,z,v,w)$, $B_i(y,z,v,w)$ $(i = 1,2,3,4)$ and $D(y,z,v,w)$ such that

$$
\begin{aligned}
F(x,y,z) &= A(y,z,v,w)S(x)+B_1(y,z,v,w)x+D(y,z,v,w),\\
F\left(u,\frac{v}{1-y},\frac{w}{1-z}\right) &= A(y,z,v,w)S(u)+B_2(y,z,v,w)u+\\
&\quad + B_1(y,z,v,w)-B_4(y,z,v,w),\\
F(x,v,w) &= A(y,z,v,w)S(x)+B_3(y,z,v,w)x+\\
&\quad +B_1(y,z,v,w)+B_2(y,z,v,w)-B_3(y,z,v,w)-\\
&\quad -B_4(y,z,v,w)+D(y,z,v,w),\\
F\left(u,\frac{y}{1-v},\frac{z}{1-w}\right) &= A(y,z,v,w)S(u)+B_4(y,z,v,w)u+\\
&\quad +B_3(y,z,v,w)-B_2(y,z,v,w)
\end{aligned}
\tag{7}
$$

for all $x \in I_1$ and $u \in I$. The functions A, B_i and D give F consistently through (7) only if they satisfy

$$
\begin{aligned}
A(y,z,v,w) &= \text{constant} =: A \text{ say},\\
B_1(y,z,v,w) &= \text{a function of } y,z =: B(y,z) \text{ say},\\
D(y,z,v,w) &= \text{a function of } y,z =: C(y,z) \text{ say},
\end{aligned}
\tag{8}
$$

where B and C are Lebesgue measurable in each variable and satisfy the functional equations

$$B(y,z)-B\left(\frac{y}{1-v},\frac{z}{1-w}\right) = C\left(\frac{v}{1-y},\frac{w}{1-z}\right) \tag{9}$$

and

$$C\left(\frac{v}{1-y},\frac{w}{1-z}\right)-C\left(\frac{y}{1-v},\frac{z}{1-w}\right)+C(y,z) = C(v,w) \tag{10}$$

for every $y,z,v,w \in I^0$ with $y+v$, $z+w \in I^0$. Furthermore F is given by

$$F(x,y,z) = AS(x)+B(y,z)x+C(y,z). \tag{11}$$

Thus we obtain the following:

LEMMA 1. *If F satisfies the hypothesis in the theorem, then F has necessarily the form* (11), *where S is the Shannon function, B and C are measurable in each variable and satisfying the functional equations* (9) *and* (10).

Now we proceed to solve equations (9) and (10) for measurable solutions. We first consider equation (10) and introduce the following lemma.

LEMMA 2. *If $a_i : I^0 \to R$ $(i = 1, 2, 3, 4)$ are measurable functions satisfying the functional equation*

$$a_1(v) + a_2\left(\frac{y}{1-v}\right) = a_3(y) + a_4\left(\frac{v}{1-y}\right) \tag{12}$$

for all $y, v \in I^0$ with $y + v \in I^0$, then a_i's have derivatives of all orders on I^0.

Proof. As a first step, we shall prove that the a_i's are locally integrable (Lebesgue). Let $v_0 \in I^0$ be arbitrarily fixed and we shall first show that a_1 is locally bounded at v_0 and hence locally integrable at v_0. Let $A_n := a_2^{-1}([-n, n]) \cap a_3^{-1}([-n, n]) \cap a_4^{-1}([-n, n])$. Then A_n is measurable and the measure $\mu(A_n)$ of A_n increases to $\mu([0, 1])$ as n increases to infinity. Let $\varepsilon > 0$ be arbitrarily fixed, then there exists an A_N such that $\mu(I^0 \setminus A_N) \leqslant \varepsilon$. Let $z_0 = \min\{v_0, 1-v_0\}$. It follows, by a simple computation similar to that in [3] (Lemma 3), that

$$\mu\big((1-v_0)A_N \cap A_N \cap (1 - v_0 A_N^{-1})\big) \geqslant \mu(]0, z_0[) - 3\varepsilon$$

which will be positive if we choose $\varepsilon = \frac{1}{6}\mu(]0; z_0[)$ at the beginning. Hence by the continuity of $v \to \mu\big((1-v)A_N \cap A_N \cap (1 - vA_N^{-1})\big)$ at v_0 (cf. [3], Lemma 2) there exists a neighbourhood $N(v_0)$ of v_0 such that

$$\mu\big((1-v)A_N \cap A_N \cap (1 - vA_N^{-1})\big) > 0$$

for all $v \in N(v_0)$. In particular we have

$$(1-v)A_N \cap A_N \cap (1 - vA_N^{-1}) \neq \emptyset$$

for each $v \in N(v_0)$. This is equivalent to the existence of y such that $y/(1-v)$, y, $v/(1-y)$ are all in A_N for each $v \in N(v_0)$. Thus by (12), with such choice of y,

$$|a_1(v)| = \left| -a_2\left(\frac{y}{1-v}\right) + a_3(y) + a_4\left(\frac{v}{1-y}\right) \right| \leqslant 3N$$

for each $v \in N(v_0)$. So a_1 is locally bounded and hence locally integrable at v_0.

By symmetry, a_3 is locally integrable in I^0.

To see that a_2 is locally integrable at an arbitrary point $t_0 \in I^0$, we choose $y_0, v_0 \in I^0$ to be such that $t_0 = y_0/(1-v_0)$. If we let $s_0 = v_0/(1-y_0)$, the substitution $t = y/(1-v)$ and $s_0 = v/(1-y)$ in (12), then enables us to get

$$a_2(t) = a_3\left(\frac{(1-s_0)t}{1-s_0 t}\right) + a_4(s_0) - a_1\left(\frac{(1-t)s_0}{1-s_0 t}\right) \tag{13}$$

for every t in some neighbourhood of t_0. Thus the local integrability of a_2 at t_0 follows from that of a_3 at y_0 and of a_1 at v_0.

The local integrability of a_4 follows by symmetry.

With the local integrability of the a_i's we shall prove that a_i's have derivatives of all orders. We integrate equation (12) w.r.t. y and obtain

$$(14)\quad (t-s)a_1(v) = \int_s^t a_3(y)dy + v\int_{\frac{v}{1-s}}^{\frac{v}{1-t}} \frac{a_4(z)}{z^2}dz - (1-v)\int_{\frac{s}{1-v}}^{\frac{t}{1-v}} a_2(z)dz$$

for appropriate s and t. The continuity of the right-hand side of (14) in v implies that a_1 is continuous. By symmetry, a_3 is continuous. From (13), a_2 is continuous and hence by symmetry a_4 is continuous. The continuity of a_2, a_3 and a_4 imply that the right-hand side of (14) is differentiable in v, hence a_1 is differentiable. Repeating the above arguments we get the differentiability of the a_i's of all orders.

Remark 1. The method adopted to prove Lemma 2 is quite standard (cf. [5]).

LEMMA 3. *If $a_i: I^0 \to R$ $(i = 1, 2, 3, 4)$ are measurable functions satisfying equation* (12), *then they are given by*

$$(15)\quad \begin{aligned} a_1(v) &= (C_1+C_2)\log(1-v)+C_4\log v - C_3+C_6,\\ a_2(s) &= C_1\log(1-s)+C_2\log s+C_3,\\ a_3(y) &= (C_1+C_4)\log(1-y)+C_2\log y - C_5+C_6,\\ a_4(t) &= C_1\log(1-t)+C_4\log t+C_5 \end{aligned}$$

on I^0, where C_i $(i = 1, 2, \ldots, 6)$ are constants.

Proof. We differentiate (12) w.r.t. v and then the resultant w.r.t. y and get

$$\frac{1}{(1-v)^2}a_2'\left(\frac{y}{1-v}\right)+\frac{y}{(1-v)^3}a_2''\left(\frac{y}{1-v}\right) = \frac{1}{(1-y)^2}a_4'\left(\frac{v}{1-y}\right)+\frac{v}{(1-y)^3}a_4''\left(\frac{v}{1-y}\right)$$

which, by the substitution $y/(1-v) = s$ and $v/(1-y) = t$, gives

$$\frac{(ts-1)^2}{(t-1)^2}a_2'(s)+\frac{s(ts-1)^2}{(t-1)^2}a_2''(s) = \frac{(ts-1)^2}{(s-1)^2}a_4'(t)+\frac{t(ts-1)^2}{(s-1)^2}a_4''(t)$$

and may be separated as

$$(s-1)^2[a_2'(s)+sa_2''(s)] = (t-1)^2[a_4'(t)+ta_4''(t)].$$

Thus

$$(s-1)^2[a_2'(s)+sa_2''(s)] = C_1,$$

where C_1 is a constant. The solution of the above differential equation is

(16) $$a_2(s) = C_1\log(1-s)+C_2\log s+C_3.$$

Similarly for a_4 we have

(17) $$a_4(t) = C_1\log(1-t)+C_4\log t+C_5.$$

A substitution of (16) and (17) into (12) gives

(18) $$\begin{aligned} a_1(v) &= (C_1+C_2)\log(1-v)+C_4\log v-C_3+C_6,\\ a_3(y) &= (C_1+C_4)\log(1-y)+C_2\log y-C_5+C_6 \end{aligned}$$

and this completes the proof of Lemma 3.

LEMMA 4. *If C is measurable in each variable and satisfies* (10), *then it is given by*

(19) $$C(s,t) = d_1\log(1-s)+e_1\log(1-t)+e_2$$

for all $s, t \in I^0$ with constants d_1, e_1 and e_2.

Proof. For each fixed $w, z \in I^0$ with $w+z \in I^0$ equation (10) reduces to the form of (12) and hence by Lemma 3 there exist constants $C_i(w,z)$, $i = 1, 2, \ldots, 6$, such that

(20) $$\begin{aligned} C(v,w) &= \big(C_1(w,z)+C_2(w,z)\big)\log(1-v)+C_4(w,z)\log v-\\ &\quad -C_3(w,z)+C_6(w,z),\\ C\left(s,\frac{z}{1-w}\right) &= C_1(w,z)\log(1-s)+C_2(w,z)\log s+C_3(w,z),\\ C(y,z) &= \big(C_1(w,z)+C_4(w,z)\big)\log(1-y)+C_2(w,z)\log y-\\ &\quad -C_5(w,z)+C_6(w,z),\\ C\left(t,\frac{w}{1-z}\right) &= C_1(w,z)\log(1-t)+C_4(w,z)\log t+C_5(w,z). \end{aligned}$$

The four equations in (20) give C consistently only if C has the form

(21) $$C(v,w) = d_1\log(1-v)+d_2(w)$$

on $I^0\times I^0$, where d_2 is a measurable function satisfying the functional equation

(22) $$d_2(w)+d_2\left(\frac{z}{1-w}\right) = d_2(z)+d_2\left(\frac{w}{1-z}\right)$$

for all $w, z \in I^0$ with $w+z \in I^0$. By Lemma 3 again, the measurable solutions d_2 of (22) are given by

$$d_2(w) = e_1\log(1-w)+e_2$$

on I^0, where e_1 and e_2 are constants. Hence we may write (21) as (19) and the lemma is proved.

With C given by (19), equation (9) reduces to

$$(23)\quad B(y,z)-B\left(\frac{y}{1-v},\frac{z}{1-w}\right) = d_1\log\left(1-\frac{v}{1-y}\right)+e_1\log\left(1-\frac{w}{1-z}\right)+e_2$$

for all $y, z, v, w \in I^0$ with $y+v,\ z+w \in I^0$. Letting $1-v = 2y$ and $1-w = 2z$ for $y, z \in]0, \frac{1}{2}[$ in (23), we see that B has the form

$$(24)\quad B(y,z) = d_1\log\frac{y}{1-y}+e_1\log\frac{z}{1-z}+e_3$$

on $]0, \frac{1}{2}[\times]0, \frac{1}{2}[$, where d_1, e_1 and e_3 are constants. Setting $v = w = \frac{1}{2}$ in (23) we get

$$(25)\quad B(y,z)-B(2y,2z) = d_1\log\left(1-\frac{1}{2(1-y)}\right)+e_1\log\left(1-\frac{1}{2(1-z)}\right)+e_2$$

for all $y, z \in]0, \frac{1}{2}[$. From (24) and (25) we get

$$(26)\quad B(2y,2z) = d_1\log\left(\frac{2y}{1-2y}\right)+e_1\log\left(\frac{2z}{1-2z}\right)+e_4$$

for all $y, z \in]0, \frac{1}{2}[$ with constants d_1, e_1 and e_4. This gives the following lemma.

LEMMA 5. *The solution B of* (9) *with C given by* (19) *is*

$$(27)\quad B(y,z) = d_1\log\left(\frac{y}{1-y}\right)+e_1\log\left(\frac{z}{1-z}\right)+e_4$$

and consequently $e_2 = 0$.

Combining Lemmas 1, 4 and 5 we have from (11), (19) and (27) that F is of the form

$$F(x,y,z) = AS(x)+\left[d_1\log\left(\frac{y}{1-y}\right)+e_1\log\left(\frac{z}{1-z}\right)+e_4\right]x+ + d_1\log(1-y)+e_1\log(1-z)$$

and this completes the proof of the theorem.

Remark 2. Equations analogous to (3) when F is an n-place function can be solved along similar lines as one goes from $n = 2$ to $n = 3$.

Remark 3. The measurable solutions of (3) are hence differentiable.

References

[1] J. Aczél and P. Nath, *Axiomatic characterizations of some measures of divergence in information*, Z. Wahr. Verw. Geb. 21 (1972), p. 215–224.
[2] P. Kannappan, *On various characterizations of generalized directed divergence* (submitted).
[3] — and C. T. Ng, *Measurable solutions of functional equations related to information theory*, Proc. Amer. Math. Soc. 38 (1973), p. 303–310.
[4] — and P. N. Rathie, *An axiomatic characterization of generalized directed divergence*, Kybernatika 9 (1973), p. 330–337.
[5] P. M. Lee, *On the axioms of information theory*, Ann. Math. Statist. 35 (1964), p. 414–418.
[6] P. Nath, *On the measures of errors in information*, J. Math. Sci. Delhi 3 (1968), p. 1–16.

UNIVERSITY OF WATERLOO
WATERLOO, ONTARIO

Reçu par la Rédaction le 20. 12. 1972